

## Többnyelvű közmondás-adatbázis

Hrisztova-Gotthardt Hrisztalina<sup>1</sup>

Pécsi Tudományegyetem  
Bölcsészettudományi Kar, Nyelvtudományi Doktori Iskola  
Postai cím: 7632 Pécs, Olga u.1.X.32  
E-mail: xpucu@freeweb.hu

**Kivonat** A cikk egy többnyelvű közmondás-adatbázist mutat be. A hálószerű séma szerint felépülő WEB-alapú adatbázis részletesen prezentálja a benne szereplő közmondások jegyeit és a hozzájuk tartozó fontos információkat, valamint lehetővé teszi a különböző nyelvekben levő közmondások közötti szemantikai kapcsolatok létrehozását. Az adatbázis általános hozzáférhetősége, részletes keresés, kiegészítés és javítás lehetősége segítheti a nyelvész- és folklór-kutatók, a tankönyvkészítők, a fordítók és a nyelvtanárok munkáját, akik tudományos, szerkesztői, fordítási és egyéb munkájuk során korpuszként, illetve gyors tájékozódáshoz és ellenőrzéshez használhatják.

### 1 Bevezetés

A disszertációs projektem keretében, amely bolgár, magyar és német közmondások kontrasztív elemzését a világ nyelvi képe fényében tűzi ki célul, felmerült a kutatásom során használható adatbázis hiányának problémája. Munkám során olyan hipotéziseket kívánok bizonyítani, amelyekhez egy ilyen adatbázis megléte elengedhetetlen. Jelenleg azonban még nem létezik számítógépes közmondás-adatbázis, amely a többnyelvűséget biztosítja. A kétnyelvű szótárak is – kétnyelvű adatforrásként – e téren hiányosnak bizonyulnak; a bennük szereplő parómiák száma igen csekély, a megfelelések nem ritkán pontatlanok vagy számuk nem teljes. A többnyelvű közmondás-gyűjtemények gyakran csak néhány – bizonyos szempontok alapján kiválasztott – nyelvet vesznek alapul, és ilyen módon leszűkítik azt a lehetőséget, hogy különböző nyelvi csoportokba és családokba tartozó nyelveket hasonlítsunk össze. Nikolaj Ikonomov (bolgár, albán, görög, román, orosz, szerb és török közmondások) és Szergej Vlahov (bolgár, angol, francia, német és latin közmondások) gyűjteményei többnyelvűsége törekednek, mégis egy központi nyelv – a bolgár – köré csoportosulnak, amelyet vagy csak a szomszédos, vagy csak a „világnyelvekkel” ütköztetnek. A fentiek értelmében e projekt során egy többnyelvű közmondás-adatbázis kidolgozása a célom, melyet a következőkben ismertetek.

## 2 Adatbázis: célok és követelmények

Egy leendő számítógépes közmondás-adatbázis több célt is szolgálhat. Korpuszként elsősorban kutatási célokra használható. Ahogyan Schneider is említi, egy jól strukturált és kidolgozott korpusz nagy segítséget nyújthat, és számos alapot biztosíthat hipotézisek bizonyítása, a nyelv részletes, empirikus-tudományos leírása során (Schneider 2000)[6]. Másodsorban egy többnyelvű adatbázis fontos szerepet játszhat az anyanyelv és idegen nyelv oktatásában – tankönyv és tanítási anyagok készítői információforrásként használhatják, amely az adott nyelv legelterjedtebb közmondásait és más nyelvű megfeleléseit tartalmazza. Nem utolsó sorban hasznát vehetik a különböző szótárak szerzői, attól függetlenül, hogy egy-, két-, vagy többnyelvű szótárról van-e szó. Ebben az értelemben a fordítók is kihasználhatják az előnyeit, mivel a célnyelvben nem mindig áll rendelkezésükre „szó-szerinti”, teljes azonosságot mutató parömia.

Az adatbázis-követelményeket illetően, a következőket tűztém ki legfontosabbnak és elengedhetetlennek:

- Az adatbázis struktúrája legyen átlátható
- Sokoldalúan legyen lekérdezhető
- Legyen aktualizálható: adjon lehetőséget törlésre, kiegészítésre, javításra
- Az általános hozzáférhetőség miatt legyen WEB-alapú

Dobrovolskij szerint egy jó adatbázist nem csak a reprezentatív adatok elegendő mennyisége tesz ki, hanem az egyes adatok részletes leírása (Dobrovolskij 2002:430)[3]. Emiatt a fenti négy pont kiegészül az alábbi ötödikkel:

- Az adatbázis az adatok (a közmondások) lehetőség szerinti legszélesebb körű leírását tartalmazza

## 3 Az adatok

Mivel a disszertációs projektem keretében a három – bolgár, magyar és német – nyelvben szereplő lehetőleg legismertebb és legelterjedtebb közmondásokat szándékozom elemezni és összevetni, ennek létrehozásának első szakaszában az úgynevezett „parömiológiai minimumokkal” fogok dolgozni. A három elemzendő nyelv közül csak a magyar rendelkezik kidolgozott parömiológiai minimummal, amely 158, az adatközlők legalább 90%-a által ismert közmondást foglal magában (ld. Tóthné Litovkina, Anna 1996)[7]. A magyar közmondás-minimumot alapul véve, felépül a leendő adatbázis, és később kiegészül német és bolgár nyelvű adatokkal. A közmondások várható száma a kevésbé ismertekkel együtt több ezerre tehető nyelvenként.

## 4 Adatstruktúra

### 4.1 Szemantikai rész

Az általam folytatott elemzés fő tárgya a három nyelv közmondásaiban kibontakozó világ nyelvi képének feltárása, illetve rekonstruálása. Ezen belül különös figyelmet fordítok a teljes azonosságra, az ekvivalens formák hiányára, és a gondolatbeli megfelelés, de ugyanúgy a felszíni (képbeli) eltérés bemutatására. A kognitív nyelvészet téziseire hivatkozva a közmondásokat a világról szóló ítéletekként definiálhatjuk (Bańczerowski a,b)[1][2]. Így az analízis során három alapvető kérdésre keresünk választ:

- Miről léteznek ítéletek az adott nyelvben?
- Mi egy-egy ítélet mondanivalója?
- Hogyan (milyen nyelvi kép segítségével) juttatja kifejezésre a három nyelvközösség az ítéletet?

A kérdéseket csak fordított sorrendben tudjuk megválaszolni – a felszíni struktúrától kiindulva (a nyelvi megvalósítás) a mély struktúra (kognitív szint) felé haladva. A legtöbb különbség a felszínen várható, mivel a három nyelvközösség a történeti, társadalmi, kulturális és nyelvi fejlődésénél és rendszerénél fogva más-más képet vesz igénybe az adott ítélet megfogalmazására. E hipotézis bizonyítására mindhárom nyelvből a parömiológiai minimumba tartozó közmondásokat szemantikai kategóriák útján visszavezetjük kognitív eredetükhöz, amely magában foglalja a fizikális tapasztalatok és mentális élmények által a világról megszerzett tudást és ennek a tudásnak a nyelvben történő kifejeződését. A visszavezetés struktúrája a következőképpen ábrázolható: (ábra 1).

Az adatbázis struktúra reprezentációs módja matematikailag egy irányított, véges, egyszerű gráf. A gráf ábráján a legalsó szinten találhatók a konkrét közmondások, ezek közvetlen elődei pedig az adott közmondás egyes jelentései. A gráfról leolvasható a szemantikailag rokon közmondások relációja, azaz ha két közmondás közötti útvonal „elég rövid”, akkor azok jelentései között szoros rokonság áll fenn.

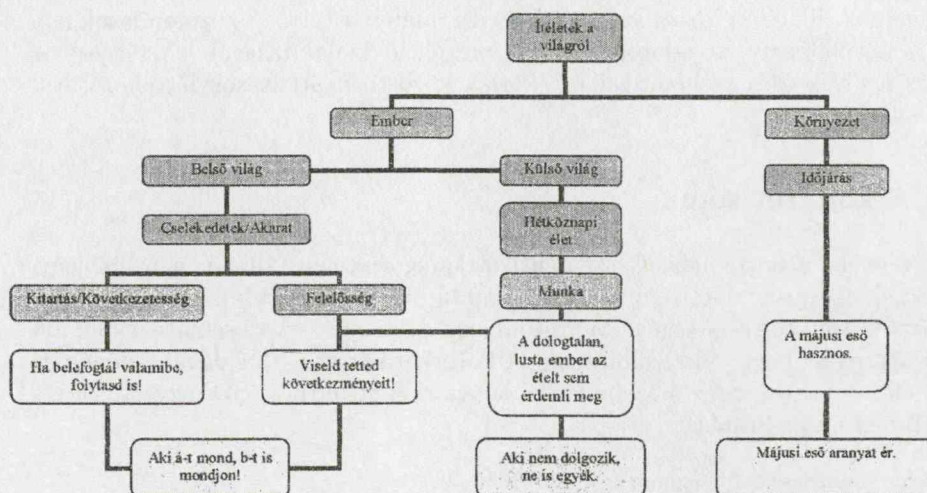
A fentiekben bemutatott módszer szükségessé teszi az ítéletek fogalmi kategóriákra való bontását, a közmondások kategóriákba való besorolását, és az adatbázis szemantikai részének létrehozását. Az így létrehozott „szemantikai” közmondás-adatbázis szolgál alapul a további elemzésekhez.

### 4.2 Közmondások eredetéről szóló információk

Írásbeli forrásokban fellelhető eredet is elengedhetetlen egy közmondás pontos leírásánál. Latin, középkorból származó, bibliai stb. eredetű parömiáknál fontos ennek az információnak a megadása.

### 4.3 Forrás

Közmondás-gyűjteményekből, különböző írásbeli vagy szóbeli szövegekből vett parömiáknál leírandó(k) a pontos forrás(ok).



1. ábra. A közmondások szemantikai rekonstrukciója (Az XML adatbázis egyszerűsített vizualizált kivonata)

#### 4.4 Minősítések (stílus, beszédszándék, kor, tájnyelv)

Az adatok részletes leírásánál nélkülözhetetlen ezen jegyeknek a jelölése, például: *irod.* stílusértékként, *táj.* mint annak a jelölése, hogy egyes közmondások csak bizonyos régiókban ismertek, *rég.* mint annak a jele, hogy azok csak régi forrásokban találhatók stb.

#### 4.5 Formai jegyek

Nem ritka jelenség az, hogy egy közmondásnak több formai változata/variánsa is létezik. A variánsokat a szótárak és a gyűjtemények szögletes zárójellel jelölik: *Embert szaván, ökröt szarván. [Ökröt szarvánál, embert szavánál.]* (O. NAGY 1999:504)[5] Hasonlóképpen sok közmondásnak létezik hosszabb és rövidebb alakja is. Az a rész, amely nem tartozik szükségszerűen a közmondáshoz, a gyűjteményekben kerek zárójelek között szerepel: *(A) szegény embert (még) az ág is húzza.* (O. NAGY 1999:115)[5] Számos közmondásnál mind a két lehetőséggel számolni kell. Ezek is abszolút módon szükségesek az adatok teljes leírásához.

### 5 Nyelvek közti megfeleltetések

Az eddigiekben egy-egy nyelv közmondásainak reprezentációját ismertettünk, azonban olyan adatbázis-struktúrára van szükség, amely több nyelvből származó megfelelő parómiákat kapcsol össze. Az egyes adatbázis-lekérdezések

eredményeképpen láthatóvá kell válnia a teljesen vagy részben azonos formáknak, illetve a közös szemantikai tartományba tartozó közmondásoknak. Ez a követelmény az adatbázis séma megfelelő kialakításával lehetséges: az egyes nyelvek és a szemantikai kategóriák közötti hivatkozások létrehozásával teljesíthető.

## 6 A konkrét séma

Nyelvenként a közmondások szemantikai kategóriái egy fájlban, a többi jegy, érték, ill. az egyéb adatok egy másik fájlban tárolódnak. A két fájl közötti kapcsolatokat azonosítók segítségével valósítom meg. Erre a hálószerű sémára leginkább az XML nyelv bizonyul alkalmasnak. A következő ábra egy példával szemlélteti az ítélet-doménnek, azaz a szemantikai kategóriák struktúrájának egyszerűsített XML-beli megvalósítását.

```
<? xml version="1.0" encoding="ISO-8859-2" ?>
<ítélet>Ítéletek a világról
  <ítélet>Ember
    <ítélet>Belső világ
      <ítélet>Cselekedetek/Akarat
        <ítélet>Kitartás/Következetesség
          <jelentés id="1">Ha belefogtál valamibe, folytasd is!</jelentés>
        </ítélet>
        <ítélet>Felelősség
          <jelentés id="2">Viseld tetteid következményeit!</jelentés>
        </ítélet>
      </ítélet>
    </ítélet>
  <ítélet>Külső világ
    <ítélet>Hétköznapi élet
      <ítélet>Munka
        <jelentés id="3">A dologtalan, lusta ember az ételt sem érdemli meg
        </jelentés>
      </ítélet>
    </ítélet>
  </ítélet>
</ítélet>
<ítélet>Környezet
  <ítélet>Időjárás
    <jelentés id="4">A májusi eső hasznos.</jelentés>
  </ítélet>
</ítélet>
</ítélet>
```

Megfelelő lekérdező felülettel kiegészítve az XML teljes mértékben ellátja és kielégíti a második pontban bevezetett ismérveket. Ennek megvalósításához természetesen szükséges valamilyen XML-t ismerő szkriptnyelv (PHP vagy PERL), amely az adatbázist webszerveren keresztül összekapcsolja a külvilággal.

Így az adatfeltöltéshez a feltöltő személyeknek nem szükséges ismerniük az XML nyelvet, sem egyéb formai vagy programozási technikákat. A megjelenítő, lekérdező és adatbeviteli felület teljes egészében a HTML nyelv szabványaira épül, így tetszőleges böngészőprogrammal az említett műveletek könnyedén elvégezhetők. Emiatt bárki korlátozás nélkül kapcsolódni tud a projekthez a világhálón keresztül akár a három kiinduló nyelven túl esetleges további nyelvek bevonásával is.

## 7 Záró gondolatok

A jelenlegi adatbázis-projekt egy komoly hiányt kíván pótolni, amelyre eddig csak részleges megoldás létezik, nevezetesen ilyen a *The Matti Kuusi international type system of proverbs*[4]. Outi Lauhakangas dolgozta ki a Matti Kuusi által összeállított közmondás-gyűjtemény internetes változatát, amely kizárólag forrás-megjelölést és kategóriákba sorolást nyújt, valamint csak keresésre ad lehetőséget a kutatók számára. A bemutatott közmondás-adatbázis a nyelvi adatokkal (a konkrét esetben a parómiákkal) kapcsolatos teljes spektrumot igyekszik felölelni, és általánosan hozzáférhető, bővíthető korpuszként rendelkezésre áll a kutatók (első sorban nyelvészek és folkloristák) számára.

## Hivatkozások

1. Bańcerowski Janusz (a): *A kognitív nyelvészet alapelvei*. In: Magyar Nyelvőr online <http://www.c3.hu/nyelvor/period/1231/123107.htm>.
2. Bańcerowski Janusz (b): *A világ nyelvi képe mint a szemantikai kutatások tárgya*. In: Magyar Nyelv online: <http://www.c3.hu/magyar nyelv/99-2/banczer.html>.
3. Dobrovolskij, Dmitrij (2002): *Phraseologie als Datenbank*. In: Hartmann, Dietrich/Wirren, Jan (Hrsg.): *Wer Ä sagt, muss auch B sagen. Beiträge zur Phraseologie und Sprichworforschung aus dem Westfälischen Arbeitskreis*. Hohengehren: Schneider Verlag. pp. 429-432.
4. Lauhakangas, Outi: *The Matti Kuusi international type system of proverbs*. In: <http://lauhakan.home.cern.ch/lauhakan/cerp.html>
5. O. Nagy Gábor (1999): *Magyar szókészletek és közmondások*. Talentum Kiadó.
6. Schneider, Gerold (2000): *Korpuszlinguistik I. Morphologieanalyse und Lexikonaufbau*. In: <http://www.ifi.unizh.ch/CL/gschneid/LexMorphVorl/Lexikon08.Corpora1.html>
7. Tóthné Litovkina, Anna (1996): *Parómiológiai felmérés Magyarországon (Milyen formában és változatban élnek a legismertebb közmondások, és mi határozza meg az ismeretüket?)*. In: Magyar Nyelv 4. pp. 439-458.
8. Zunker, Giesela/Rapp Reinhard (1994): *Maschinenlesbare deutsch- und englischsprachige Textkorpora*. In: Hallwachs, D.W./Stütz, I. (Hrsg.). *Sprache - Sprechen - Handeln*. Akten des 28. Linguistischen Kolloquiums, Graz 1993, Band 2. Tübingen: Niemeyer, pp. 341-348.
9. Влахов, Сергей: *Съпоставителен речник на пословици*. - София: Издателство ЕТО 1998
10. Икономов, Николай: *Балканска народна мъдрост*. - София: Издателство на Българската академия на науките 1968